October 22, 1997

# Trend Adjustment Prior to Testing for the Cointegrating Rank of a VAR Process*

by

Pentti Saikkonen and Helmut Lütkepohl

Department of Statistics
University of Helsinki
P.O. Box 54
SF-00014 University of Helsinki
FINLAND

Tel.: +358-9-1918867
Fax: +358-9-1918872

Institut für Statistik und Ökonometrie
Wirtschaftswissenschaftliche Fakultät
Humboldt–University
Spandauer Str. 1
10178 Berlin
GERMANY
Tel.: +49-30-2093-5718
Fax: +49-30-2093-5712

## Abstract

Testing the cointegrating rank of a vector autoregressive process which may have a deterministic linear trend is considered. Previous proposals for dealing with such a situation are either to allow for a deterministic trend term in computing a suitable test statistic or else remove the linear trend first and then derive the test statistic from the trend-adjusted data. In this study the latter approach is considered and a new, simple method for trend removal is proposed which is based on estimating the trend parameters under the null hypothesis. LR (likelihood ratio) and LM (Lagrange multiplier) type test statistics are derived on the basis of the trend-adjusted data and their asymptotic distributions are considered under the null hypothesis and under local alternatives. A simulation comparison with other proposals is performed which demonstrates the potentially superior small sample performance of the new tests.

*We are grateful to the Deutsche Forschungsgemeinschaft, SFB 373, for financial support and we thank Kirstin Hubrich and Ralf Brüggemann for performing the computations.

# 1 Introduction

Trending behaviour is an obvious feature of many time series. Sometimes it is not clear from the outset whether a trend is best modeled as a deterministic polynomial function or as a random component induced by unit roots in a stochastic process. In many situations it is necessary to allow for both components because there is no prior knowledge on the type of trend. These considerations have led to the development of unit root tests in the presence of polynomial trends. There are also proposals how to allow for polynomial trends in testing for the number of cointegrating relationships in a vector autoregressive (VAR) process. A seemingly simple way to do so is to include an intercept term in the VAR process. Such a drift term may induce a linear trend in the data series. Unfortunately, the popular LR (likelihood ratio) tests for the cointegrating rank proposed by Johansen (1991, 1995) are not similar in this case, that is, the asymptotic null distribution depends on whether there actually is a deterministic trend term or not. Therefore, Perron & Campbell (1993), Rahbek (1994) and Johansen (1992, 1995) propose to include linear trend terms in the model and thereby construct similar tests.

In contrast to these proposals, Lütkepohl & Saikkonen (1997) (henceforth L&S) suggest to subtract the deterministic trend in a first step and then apply LM (Lagrange multiplier) type tests for the cointegrating rank. Subtracting the trend first is also suggested by Stock & Watson (1988) in the context of their cointegration tests. Their proposal for trend removal is different from that of L&S, however. In this study we will build on the latter paper and propose an alternative method for trend-adjustment where the trend parameters are estimated under the null hypothesis of the cointegrating rank being $r_0$, say. For this purpose a simple GLS (generalized least squares) method is developed for estimating the trend parameters. These estimators are then used for removing the linear trend from the data and both the asymptotic and small sample properties of LM and LR type tests based on trend-adjusted data will be explored. It turns out that these tests have favorable properties compared to LR tests of the type proposed by Johansen and Perron & Campbell for processes with deterministic trends.

The structure of the paper is as follows. In the next section the framwork of the analysis is laid out. Estimation of the parameters of the deterministic trend term is considered in Section 3. LM and LR type tests for the number of cointegration relations based on the trend-adjusted series are discussed in Section 4. Small sample properties of the new tests are investigated in Section 5 by means of a small simulation experiment. Conclusions are given in Section 6 and proofs are presented in an appendix.

The following notation is used throughout. The symbol $y_t = (y_{1t}, \ldots, y_{nt})'$ is reserved for an $n$-dimensional vector of observable time series variables. The lag and differencing operators are denoted by $L$ and $\Delta$, respectively, that is, $Ly_t = y_{t-1}$ and $\Delta y_t = y_t - y_{t-1}$. The symbol $I(d)$ is used to denote a process which is integrated of order $d$, that is, it is stationary or asymptotically stationary after differencing $d$ times while it is still nonstationary after differencing just $d-1$ times. $\mathbf{B}$ denotes a multivariate standard Brownian motion of suitable dimension. The symbol $\xrightarrow{d}$ signifies convergence in distribution. $\lambda_{max}(A)$, $\mathrm{tr}(A)$ and $\mathrm{rk}(A)$ denote the maximal eigenvalue, the trace and the rank of the matrix $A$, respectively. If $A$ is an $(n \times m)$ matrix of full column rank $(n > m)$ we denote its orthogonal complement by $A_\perp$. In other words, $A_\perp$ is an $(n \times (n-m))$ matrix of full column rank and such that $A'A_\perp = 0$. The orthogonal complement of a nonsingular square matrix is zero and the orthogonal complement of zero is an identity matrix of suitable dimension. An $(n \times n)$ identity matrix is denoted by $I_n$. LS and GLS are used to abbreviate least squares and generalized least squares, respectively, and DGP is short for data generation process. As a general convention, a sum is defined to be zero if the lower bound of the summation index exceeds the upper bound.

# 2 The Model Framework

Consider the DGP of an $n$-dimensional multiple time series $y_t = (y_{1t}, \ldots, y_{nt})'$ defined by

$$y_t = \mu_0 + \mu_1 t + x_t, \qquad t = 1, 2 \ldots, \tag{2.1}$$

where $\mu_0$ and $\mu_1$ are unknown, fixed $(n \times 1)$ parameter vectors and $x_t$ is an unobservable error process with VAR representation of order $p$ (VAR($p$)):

$$x_t = A_1 x_{t-1} + \cdots + A_p x_{t-p} + \varepsilon_t. \tag{2.2}$$

Here the $A_j$ are $(n \times n)$ coefficient matrices. Subtracting $x_{t-1}$ on both sides of (2.2) and rearranging terms gives the error correction (EC) form

$$\Delta x_t = \Pi x_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta x_{t-j} + \varepsilon_t, \qquad t = p+1, p+2, \ldots, \tag{2.3}$$

where $\Pi = -(I_n - A_1 - \cdots - A_p)$ and $\Gamma_j = -(A_{j+1} + \cdots + A_p)$ $(j = 1, \ldots, p-1)$ are $(n \times n)$. We assume that the error term $\varepsilon_t$ is a martingale difference sequence such that $E(\varepsilon_t|\varepsilon_{t-1}, \ldots, \varepsilon_1) = 0$, $E(\varepsilon_t \varepsilon_t'|\varepsilon_{t-1}, \ldots, \varepsilon_1) = \Omega$ is a nonrandom positive definite matrix and the fourth moments are bounded. Moreover, for convenience, we impose the initial value condition $x_t = 0$, $t \leq 0$. Our results remain valid if the initial values have some fixed distribution which does not depend on the sample size.

We assume that the process $x_t$ is at most $I(1)$ and cointegrated with cointegrating rank $r$. Hence, the matrix $\Pi$ can be decomposed as

$$\Pi = \alpha \beta', \tag{2.4}$$

where $\alpha$ and $\beta$ are $(n \times r)$ matrices of full column rank. Note that we do not exclude the possibility that $x_t$ is $I(0)$, that is, the cointegrating rank may be $n$. Under these assumptions $\beta' x_t$ is a zero mean (asymptotically) stationary process (see Engle & Granger (1987) and Johansen (1991)). Defining

$$\Psi = I_n - \Gamma_1 - \cdots - \Gamma_{p-1} = I_n + \sum_{j=1}^{p-1} j A_{j+1}$$

it follows from Johansen's (1991) formulation of Granger's representation theorem that

$$x_t = C \sum_{i=1}^{t} \varepsilon_i + \xi_t, \qquad t = 1, 2, \ldots, \tag{2.5}$$

where, apart from the specification of initial values, $\xi_t$ is a stationary process and $C = \beta_\perp (\alpha_\perp' \Psi \beta_\perp)^{-1} \alpha_\perp'$.

In the DGP (2.1) the deterministic trend is added to the stochastic part. One advantage of this formulation is that the trend is clearly seen to be at most linear. It follows from (2.1)/(2.2) that $y_t$ also has a VAR($p$) representation

$$y_t = \nu_0 + \nu_1 t + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t, \qquad t = p+1, p+2, \ldots, \tag{2.6}$$

where

$$\nu_0 = -\Pi \mu_0 + (\Psi - I_n)\mu_1 \quad \text{and} \quad \nu_1 = -\Pi \mu_1 \tag{2.7}$$

(see L&S). The corresponding EC form is

$$\Delta y_t = \nu_0 + \nu_1 t + \Pi y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \varepsilon_t, \qquad t = p+1, p+2, \ldots, \tag{2.8}$$

or

$$\Delta y_t = \nu + \alpha(\beta' y_{t-1} - \tau(t-1)) + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \varepsilon_t, \qquad t = p+1, p+2, \ldots, \tag{2.9}$$

where $\nu = -\Pi\mu_0 + \Psi\mu_1$ and $\tau = \beta'\mu_1$.

In this framework we are interested in testing

$$H(r_0) : \mathrm{rk}(\Pi) = r_0 \qquad \text{vs.} \qquad \bar{H}(r_0) : \mathrm{rk}(\Pi) > r_0, \tag{2.10}$$

that is, the cointegrating rank being $r_0$ is tested against a rank greater than $r_0$. We will also consider tests for pairs of hypotheses

$$H(r_0) : \mathrm{rk}(\Pi) = r_0 \qquad \text{vs.} \qquad H(r_0+1) : \mathrm{rk}(\Pi) = r_0 + 1. \tag{2.11}$$

In the next section estimators of the trend parameters $\mu_0$ and $\mu_1$ will be given. These will then be used for trend-adjusting $y_t$ before tests of (2.10) or (2.11) are applied.

# 3 Estimating the Trend Parameters

In the following it is assumed that $\alpha$ and $\beta$ are $(n \times r_0)$ matrices, that is, their column dimension corresponds to the rank under the null hypothesis. The idea underlying our estimation method for the trend parameters $\mu_0$ and $\mu_1$ is to apply feasible GLS to the model (2.1). For this purpose we rewrite (2.1) as

$$A(L)y_t = G_t\mu_0 + H_t\mu_1 + \varepsilon_t, \tag{3.1}$$

where $A(L) = I_n - A_1L - \cdots - A_pL^p$, $y_t = 0$ for $t \leq 0$, $G_t = A(L)a_t$ and $H_t = A(L)b_t$ with

$$a_t = \begin{cases} 1 & \text{for} \quad t \geq 1 \\ 0 & \text{for} \quad t \leq 0 \end{cases}, \qquad b_t = \begin{cases} t & \text{for} \quad t \geq 1 \\ 0 & \text{for} \quad t \leq 0 \end{cases}.$$

Furthermore, we define

$$Q = [\Omega^{-1}\alpha(\alpha'\Omega^{-1}\alpha)^{-1/2} : \alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1/2}]. \tag{3.2}$$

It is straightforward to see that

$$QQ' = \Omega^{-1}\alpha(\alpha'\Omega^{-1}\alpha)^{-1}\alpha'\Omega^{-1} + \alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\alpha_\perp' = \Omega^{-1}. \tag{3.3}$$

Premultiplying (3.1) by $Q'$ thus results in a multivariate regression model with identity error covariance matrix. Thus, as in GLS estimation, we have found a transformation which results in a regression model with standard properties of the error term. Using this particular transformation will turn out to be convenient in the following. The idea is now to use a feasible version of the transformed regression model for estimating $\mu_0$ and $\mu_1$. Therefore we replace all other unknown parameters by suitable estimators.

Suitable estimators $\tilde{\alpha}, \tilde{\beta}, \tilde{\Gamma}_j$ and $\tilde{\Omega}$ may be obtained by a reduced rank regression of (2.9) (see Johansen (1995)). From these estimators the $A_j$ coefficient matrices may be estimated as follows:

$$\begin{aligned}
\tilde{A}_1 &= I_n + \tilde{\alpha}\tilde{\beta}' + \tilde{\Gamma}_1, \\
\tilde{A}_j &= \tilde{\Gamma}_j - \tilde{\Gamma}_{j-1}, \quad j = 2, \ldots, p-1, \\
\tilde{A}_p &= -\tilde{\Gamma}_{p-1}.
\end{aligned}$$

We define $\tilde{A}(L) = I_n - \tilde{A}_1 L - \cdots - \tilde{A}_p L^p$, $\tilde{G}_t = \tilde{A}(L)a_t$ and $\tilde{H}_t = \tilde{A}(L)b_t$. Moreover, we obtain $\tilde{\alpha}_\perp$ and $\tilde{\beta}_\perp$ from $\tilde{\alpha}$ and $\tilde{\beta}$, respectively, and replace $\Omega, \alpha$ and $\alpha_\perp$ in (3.2) by their estimators to get $\tilde{Q}$. Now we may estimate $\mu_0$ and $\mu_1$ by multivariate LS from the auxiliary regression model

$$\tilde{Q}'\tilde{A}(L)y_t = \tilde{Q}'\tilde{G}_t\mu_0 + \tilde{Q}'\tilde{H}_t\mu_1 + \eta_t, \qquad t = 1, \ldots, T. \tag{3.4}$$

We will denote the resulting estimators of $\mu_0$ and $\mu_1$ by $\tilde{\mu}_0$ and $\tilde{\mu}_1$, respectively. The estimators will be used in trend-adjusting the data prior to applying tests for the cointegrating rank. The following properties of these estimators are central for using them for this purpose.

**Theorem 1**

Under the conditions stated in the foregoing

$$\beta'(\tilde{\mu}_0 - \mu_0) = O_p(T^{-1/2}) \tag{3.5}$$

$$\beta_\perp'(\tilde{\mu}_0 - \mu_0) = O_p(1) \tag{3.6}$$

$$\beta'(\tilde{\mu}_1 - \mu_1) = O_p(T^{-3/2}) \tag{3.7}$$

$$\sqrt{T}\beta_\perp'(\tilde{\mu}_1 - \mu_1) \xrightarrow{d} N(0, \beta_\perp' C\Omega C'\beta_\perp) \tag{3.8}$$

and all the terms converge jointly in distribution with appropriate standardization. Here $C = \beta_\perp(\alpha_\perp'\Psi\beta_\perp)^{-1}\alpha_\perp'$ as before. □

The proof of this theorem is given in the appendix. The theorem shows that $\mu_0$ is estimated consistently in the direction of $\beta$ whereas it is not estimated consistently in the direction of $\beta_\perp$. This is not surprising because $\tilde{G}_t\tilde{\beta}_\perp(\tilde{\beta}'_\perp\tilde{\beta}_\perp)^{-1} = 0$ for $t \geq p+1$ and, hence, in the direction of $\beta_\perp$, $\mu_0$ is estimated just from the first $p$ observations regardless of the sample size. On the other hand, $\mu_1$ is consistently estimated in both directions.

# 4 Tests for the Cointegrating Rank

The idea underlying the LM type tests of L&S is to note that $\beta(\beta'\beta)^{-1}\beta' + \beta_\perp(\beta'_\perp\beta_\perp)^{-1}\beta'_\perp = I_n$ and, thus, (2.3) may be expressed as

$$\Delta x_t = \kappa u_{t-1} + \rho v_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta x_{t-j} + \varepsilon_t, \qquad t = 1, 2, \ldots, \tag{4.1}$$

where $u_t = \beta' x_t, v_t = \beta'_\perp x_t$, $\kappa = \Pi\beta(\beta'\beta)^{-1}$ and $\rho = \Pi\beta_\perp(\beta'_\perp\beta_\perp)^{-1}$. Recall that we are assuming that $\alpha$ and $\beta$ are $(n \times r_0)$ matrices. Hence, if $H(r_0)$ in (2.10) holds so that $\Pi = \alpha\beta'$, we have $\kappa = \alpha$ and $\rho = 0$. On the other hand, under the alternative, some columns of $\beta_\perp$ will be associated with cointegrating vectors so that $\rho \neq 0$. Therefore the idea is to test the restriction $\rho = 0$. Note, however, that $\rho$ is an $(n \times (n - r_0))$ dimensional matrix which is easily seen to be zero if and only if the potentially smaller $((n - r_0) \times (n - r_0))$ matrix $\rho^* := \alpha'_\perp \rho = 0$. Therefore the model (4.1) is premultiplied by $\alpha'_\perp$ and the linear hypothesis $H_0 : \rho^* = 0$ is tested in a feasible version of

$$\alpha'_\perp \Delta x_t = \kappa^* u_{t-1} + \rho^* v_{t-1} + \sum_{j=1}^{p-1} \Gamma^*_j \Delta x_{t-j} + \eta^*_t, \qquad t = 1, 2, \ldots, \tag{4.2}$$

where $\kappa^* := \alpha'_\perp \kappa$, $\Gamma^*_j = \alpha'_\perp \Gamma_j$ and $\eta^*_t = \alpha'_\perp \varepsilon_t$. Thus we have to test a set of linear restrictions in a linear model. For this purpose the three asymptotically equivalent LR, LM and Wald tests are available. The actual test statistic is determined by first obtaining estimators $\tilde{\alpha}, \tilde{\beta}, \tilde{\Gamma}_j$ and $\tilde{\Omega}$ from a reduced rank regression of (2.9) as in the previous section. Then suitable estimators of $\mu_0$ and $\mu_1$ are constructed and $y_t$ is trend-adjusted. Given the results of the previous section we propose to use $\tilde{\mu}_0$ and $\tilde{\mu}_1$ for this purpose. Hence, $\tilde{x}_t = y_t - \tilde{\mu}_0 - \tilde{\mu}_1 t$ and feasible versions of $v_t$ and $u_t$ are obtained as $\tilde{v}_t = \tilde{\beta}'_\perp \tilde{x}_t$ and $\tilde{u}_t = \tilde{\beta}' \tilde{x}_t$. This results in a test statistic

$$LM^{GLS}_*(r_0) = \operatorname{tr}\left\{\tilde{\rho}^* \tilde{M}^*_{vv\cdot\Delta X} \tilde{\rho}^{*'} (\tilde{\alpha}'_\perp \tilde{\Omega} \tilde{\alpha}_\perp)^{-1}\right\}, \tag{4.3}$$

where $\tilde{\rho}^*$ is the LS estimator of $\rho^*$ from (4.2) with $x_t$, $u_t$ and $v_t$ replaced by $\tilde{x}_t$, $\tilde{u}_t$ and $\tilde{v}_t$, respectively, $\tilde{\alpha}_\perp'\tilde{\Omega}\tilde{\alpha}_\perp$ is the residual covariance estimator of the error term in (4.2) and

$$\tilde{M}^*_{vv\cdot\Delta X} = \left[\sum_{t=p+1}^{T} \tilde{v}_{t-1}\tilde{v}_{t-1}' - \sum_{t=p+1}^{T} \tilde{v}_{t-1}\Delta\tilde{X}^{*\prime}_{t-1}\left(\sum_{t=p+1}^{T} \Delta\tilde{X}^*_{t-1}\Delta\tilde{X}^{*\prime}_{t-1}\right)^{-1} \sum_{t=p+1}^{T} \Delta\tilde{X}^*_{t-1}\tilde{v}_{t-1}'\right] \tag{4.4}$$

with

$$\Delta\tilde{X}^*_{t-1} = \begin{bmatrix} \tilde{u}_{t-1} \\ \Delta\tilde{x}_{t-1} \\ \vdots \\ \Delta\tilde{x}_{t-p+1} \end{bmatrix}.$$

We use the abbreviation LM here because the estimators used in the auxiliary model (4.2) are estimated under the null hypothesis of the cointegrating rank being $r_0$. The superscript GLS indicates that the GLS method of the previous section was used for estimating the trend parameters and the subscript corresponds to the notation used in L&S for a similar test statistic based on another trend-adjustment method.

Since under $H(r_0)$, $\kappa^* = 0$ we may estimate $\rho^*$ alternatively from the auxiliary model

$$\tilde{\alpha}_\perp'\Delta\tilde{x}_t = \rho^*\tilde{v}_{t-1} + \sum_{j=1}^{p-1}\Gamma_j^*\Delta\tilde{x}_{t-j} + \tilde{\eta}_t^*, \qquad t = p+1, p+2, \ldots.$$

Denoting the LS estimator of $\rho^*$ obtained in this way by $\hat{\rho}^*$, this results in a test statistic

$$LM^{GLS}(r_0) = \operatorname{tr}\left\{\hat{\rho}^*\tilde{M}_{vv\cdot\Delta X}\hat{\rho}^{*\prime}(\tilde{\alpha}_\perp'\tilde{\Omega}\tilde{\alpha}_\perp)^{-1}\right\}, \tag{4.5}$$

where $\tilde{M}_{vv\cdot\Delta X}$ is similar to $\tilde{M}^*_{vv\cdot\Delta X}$ in (4.4) with $\Delta\tilde{X}^*_{t-1}$ replaced by

$$\Delta\tilde{X}_{t-1} = \begin{bmatrix} \Delta\tilde{x}_{t-1} \\ \vdots \\ \Delta\tilde{x}_{t-p+1} \end{bmatrix}.$$

The asymptotic null distributions of the test statistics are given in the following theorem.

**Theorem 2**

If $H(r_0)$ in (2.10) is true,

$$LM_*^{GLS}(r_0), LM^{GLS}(r_0)$$

$$\xrightarrow{d} \mathrm{tr}\left\{\left(\int_0^1 \mathbf{B}_*(s)\mathrm{d}\mathbf{B}_*(s)'\right)'\left(\int_0^1 \mathbf{B}_*(s)\mathbf{B}_*(s)'\mathrm{d}s\right)^{-1}\left(\int_0^1 \mathbf{B}_*(s)\mathrm{d}\mathbf{B}_*(s)'\right)\right\},$$

where $\mathbf{B}_*(s) = \mathbf{B}(s) - s\mathbf{B}(1)$ is an $(n - r_0)$-dimensional Brownian bridge and $\mathrm{d}\mathbf{B}_*(s) = \mathrm{d}\mathbf{B}(s) - \mathrm{d}s\mathbf{B}(1)$. Here the integral $\int_0^1 \mathbf{B}_*(s)\mathrm{d}\mathbf{B}_*(s)'$ is a short-hand notation for $\int_0^1 \mathbf{B}(s)\mathrm{d}\mathbf{B}(s)' - \mathbf{B}(1)\int_0^1 s\mathrm{d}\mathbf{B}(s)' - \int_0^1 \mathbf{B}(s)\mathrm{d}s\mathbf{B}(1)' + \frac{1}{2}\mathbf{B}(1)\mathbf{B}(1)'$.

**Proof:** This result follows immediately from Theorem 1 using similar arguments as in the proof of Theorem 5.1 in L&S. □

Another test which may be considered in the present context is an 'LR' test based on the feasible model

$$\Delta\tilde{x}_t = \Pi\tilde{x}_{t-1} + \sum_{j=1}^{p-1}\Gamma_j\Delta\tilde{x}_{t-j} + e_t, \qquad t = p+1,\ldots,T. \tag{4.6}$$

Using the approach of Johansen (1988), the 'LR' statistic for testing (2.10) is

$$LR_{trace}^{GLS}(r_0) = \sum_{j=r_0+1}^{n} \log(1 + \bar{\lambda}_j) \tag{4.7}$$

where $\bar{\lambda}_1 \geq \cdots \geq \bar{\lambda}_n$ are the ordered generalized eigenvalues obtained as solutions of

$$\det(\bar{\Pi}\tilde{M}_T\bar{\Pi}' - \lambda\bar{\Omega}) = 0$$

where $\bar{\Pi}$ is the LS estimator of $\Pi$ obtained from (4.6), $\bar{\Omega}$ is the corresponding residual covariance matrix and

$$\tilde{M}_T = \left[\sum_{t=p+1}^{T}\tilde{x}_{t-1}\tilde{x}_{t-1}' - \sum_{t=p+1}^{T}\tilde{x}_{t-1}\Delta\tilde{X}_{t-1}'\left(\sum_{t=p+1}^{T}\Delta\tilde{X}_{t-1}\Delta\tilde{X}_{t-1}'\right)^{-1}\sum_{t=p+1}^{T}\Delta\tilde{X}_{t-1}\tilde{x}_{t-1}'\right].$$

For testing (2.11) we may use

$$LR_{max}^{GLS}(r_0) = \log(1 + \bar{\lambda}_{r_0+1}). \tag{4.8}$$

The asymptotic null distributions of these trace and maximum eigenvalue statistics are given in the next theorem.

Table 1. Percentage Points of the Distribution of

$$\lambda_{max}\left\{\left(\int_0^1 \mathbf{B}_*(s)\mathrm{d}\mathbf{B}_*(s)'\right)'\left(\int_0^1 \mathbf{B}_*(s)\mathbf{B}_*(s)'\mathrm{d}s\right)^{-1}\left(\int_0^1 \mathbf{B}_*(s)\mathrm{d}\mathbf{B}_*(s)'\right)\right\}.$$

where $\mathbf{B}_*(s)$ is a $d$-dimensional Brownian Bridge.

| Dimension | $d$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| | 90% | 5.47 | 11.51 | 17.66 | 23.64 | 29.53 |
| Percentage point | 95% | 6.87 | 13.37 | 19.72 | 26.05 | 32.07 |
| | 99% | 10.00 | 17.58 | 24.43 | 30.94 | 37.70 |

**Theorem 3**

If $H(r_0)$ is true, $LR^{GLS}_{trace}(r_0)$ has the same limiting distribution as $LM^{GLS}(r_0)$ given in Theorem 2 and

$$LR^{GLS}_{max}(r_0) \xrightarrow{d} \lambda_{max}\left\{\left(\int_0^1 \mathbf{B}_*(s)\mathrm{d}\mathbf{B}_*(s)'\right)'\left(\int_0^1 \mathbf{B}_*(s)\mathbf{B}_*(s)'\mathrm{d}s\right)^{-1}\left(\int_0^1 \mathbf{B}_*(s)\mathrm{d}\mathbf{B}_*(s)'\right)\right\}.$$

□

A proof of this theorem is given in the appendix. Critical values for the maximum eigenvalue test are presented in Table 1. They are determined by simulations in the same way as the critical values of the asymptotic distributions of the corresponding trace statistics. Details are, for instance, given in L&S.

L&S also derive the local power of their LM type tests against alternatives of the form $H_T(r_0) : \Pi = \alpha\beta' + T^{-1}\alpha_1\beta_1'$. A comparison with the local power of LR tests which allow for a linear time trend reveals that the LM type tests of L&S are considerably superior to the standard LR competitors proposed by Johansen and Perron & Campbell for some values of $\alpha_1$ and $\beta_1$. Therefore it is interesting to note that it follows from the proofs given in L&S in conjunction with Theorem 1 that $LR^{GLS}_{trace}(r_0)$, $LM^{GLS}_*(r_0)$ and $LM^{GLS}(r_0)$ all have the same local power as $LM_*(r_0)$ and $LM(r_0)$ under the conditions set forth in L&S. Hence, these tests are asymptotically equivalent even under local alternatives and they may be expected to have properties superior to the competing Johansen type LR tests in some situations in small samples. We will explore this possibility in more detail in the next section.

# 5 Small Sample Comparison of Tests

We have performed a limited simulation experiment to compare the small sample properties of the different test statistics and also to compare them to related statistics which were considered in the literature and which use alternative ways to deal with deterministic trends. In particular, we will compare the test statistics $LM^{GLS}(r_0)$, $LM_*^{GLS}(r_0)$ and $LR_{trace}^{GLS}(r_0)$ considered in Section 4. In addition we will use the trend-adjustment method proposed by L&S and include the resulting LM statistics corresponding to $LM(r_0)$ and $LM_*(r_0)$ in L&S. For clarity they will be denoted by $LM^{LS}(r_0)$ and $LM_*^{LS}(r_0)$, respectively, in the following. Moreover, we will consider the trace tests suggested by Perron & Campbell (1993) and Johansen (1995) for processes with deterministic linear trends. Perron & Campbell include a deterministic term in the EC form as in (2.8) and compute LR test statistics from that model. In other words, they do not remove the trend prior to analyzing the cointegrating rank but include the trend term in the estimation equation. The resulting statistics will be denoted by $LR_{trace}^{PC}(r_0)$. Critical values for these test statistics are taken from Table 1 of Perron & Campbell (1993). Without restrictions for $\nu_1$, the model (2.8) can in principle generate quadratic trends. Therefore, to enforce linear trends in computing the test statistics, Johansen (1995) considers the reparameterized model (2.9) and derives LR statistics based on this model. The $LR_{trace}$ statistic obtained in this way will be denoted by $LR_{trace}^{J}(r_0)$ in the following. Critical values for the corresponding test are, for instance, given in Table 15.4 of Johansen (1995).

Our simulations are based on the following bivariate process which was also used in Monte Carlo studies by Toda (1994, 1995) and L&S:

$$y_t = \begin{bmatrix} 0 \\ \delta \end{bmatrix} + \begin{bmatrix} \psi_1 & 0 \\ 0 & \psi_2 \end{bmatrix} y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \text{iid } N\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \theta \\ \theta & 1 \end{bmatrix} \right). \tag{5.1}$$

For the purposes of investigating LR tests for the cointegrating rank of a VAR(1) process, this type of process may be regarded as a "canonical form" from which other processes may be obtained by linear transformations of $y_t$ which leave the tests invariant (see Toda (1995)). For $\psi_1 = \psi_2 = 1$ a cointegrating rank of $r = 0$ is obtained. In this case the process consists of two nonstationary components with the second component having a deterministic linear trend if $\delta \neq 0$. The two components are independent for $\theta = 0$ whereas they are instantaneously correlated for $\theta \neq 0$. A cointegrating rank of $r = 1$ is obtained for $\psi_2 = 1$ and $|\psi_1| < 1$. In that case there is again a linear trend if $\delta \neq 0$. The process is $I(0)$ with $r = 2$ if both $\psi_1$ and $\psi_2$ are less than 1 in absolute value. In that case a nonzero $\delta$ cannot generate a linear trend. Therefore it will be set to zero for stationary processes.

Samples of sizes 100 and 200 plus 50 presample values starting with an initial value of zero were generated. The last presample values are used for estimation purposes so that the effective sample size $T - 1 = 100$ or 200. The number of replications is 1000. Rejection frequencies of the tests based on $LM^{GLS}(r_0)$, $LM^{LS}(r_0)$, $LM_*^{GLS}(r_0)$, $LM_*^{LS}(r_0)$, $LR_{trace}^{GLS}(r_0)$, $LR_{trace}^{LS}(r_0)$, $LR_{trace}^{J}(r_0)$ and $LR_{trace}^{PC}(r_0)$ are given in Tables 2 - 4. They are based on asymptotic critical values for a test level of 5%. The rejection frequencies are not corrected for the actual small sample sizes because these will also not be available in practice. In our opinion, comparing the power of tests which have unknown size in practice is not very useful. Therefore a minimal requirement for a test is that it observes the selected significance level at least approximately.

For a given set of parameter values and a given sample size, the results for the test statistics are based on the same generated time series. Hence the entries in the tables are not independent but can be compared directly. Still, for judging the results, it may be worth recalling that the standard error of an estimator of a true rejection probability $P$ based on 1000 replications of the experiment is $s_P = \sqrt{P(1-P)/1000}$ so that, for example, $s_{0.05} = 0.007$. It is also important to note that in the simulations the tests were not performed sequencially. Thus, the results for testing $H(1) : r = 1$ are not conditioned on the outcome of the test of $H(0) : r = 0$.

Table 2 contains results for processes with true cointegrating rank $r = 0$ ($\psi_1 = \psi_2 = 1$). It turns out that all the LM type tests are a bit conservative and reject $r = 0$ only in about 3% of the replications even for sample sizes of 200. In contrast, $LR_{trace}^{J}(r_0)$ and $LR_{trace}^{PC}(r_0)$ reject slightly more often than the nominal 5% whereas $LR_{trace}^{GLS}(r_0)$ and $LR_{trace}^{LS}(r_0)$ come closest to the ideal rejection rate of 5%. We have not given asymptotic results for testing $r = 1$ when the true rank is zero. For the presently considered process, all tests are seen to be conservative in this situation. Hence, it is not very likely that, on the basis of these tests, the process is mistakenly found to be stationary. Whether the trend parameters are estimated by the GLS method presented in Section 3 or by the method of L&S does not matter much if a true null hypothesis is tested. Hence, one may just as well use the simple

Table 2. Relative Rejection Frequencies of Test Statistics for DGP (5.1) with Cointegrating Rank $r = 0$ ($\psi_1 = \psi_2 = 1$), $\theta = 0$, $\delta = 1.0$, Nominal Significance Level 0.05.

| Test Statistic | $T-1=100$ $r_0 = 0$ | $T-1=100$ $r_0 = 1$ | $T-1=200$ $r_0 = 0$ | $T-1=200$ $r_0 = 1$ |
|---|---|---|---|---|
| $LM^{GLS}$ | 0.033 | 0.002 | 0.029 | 0.003 |
| $LM^{LS}$ | 0.033 | 0.007 | 0.029 | 0.010 |
| $LM_*^{GLS}$ | 0.033 | 0.003 | 0.029 | 0.008 |
| $LM_*^{LS}$ | 0.033 | 0.010 | 0.029 | 0.012 |
| $LR_{trace}^{GLS}$ | 0.052 | 0.002 | 0.037 | 0.004 |
| $LR_{trace}^{LS}$ | 0.052 | 0.010 | 0.037 | 0.013 |
| $LR_{trace}^{J}$ | 0.060 | 0.005 | 0.059 | 0.004 |
| $LR_{trace}^{PC}$ | 0.060 | 0.004 | 0.048 | 0.001 |

GLS method for estimating the trend parameters. It may also be worth pointing out that choosing a trend parameter $\delta = 1$ is not a severe restriction here because the test results turned out to be virtually the same for other values of $\delta$ including $\delta = 0$. The $LR^J$ and $LR^{PC}$ tests are in fact invariant to the choice of $\delta$.

In Table 3 results are given for a DGP with cointegrating rank $r = 1$, sample size $T - 1 = 100$ and two different values of the error correlation parameter $\theta$. The rejection frequencies for $r_0 = 0$ represent the power of the tests. Conservative tests may be expected to have reduced power. Therefore it is not surprising that the LM type tests are less powerful than the LR tests in this situation. For $\psi_1$ close to 1 (i.e., DGPs close to $H(0)$) the LR tests all have similar power which varies considerably with $\theta$, though. For processes far from $H(0)$ and $\theta = 0.8$, $LR_{trace}^{J}(r_0)$ and $LR_{trace}^{PC}(r_0)$ are a bit more powerful than $LR_{trace}^{GLS}(r_0)$ and $LR_{trace}^{LS}(r_0)$. Of course, in this case all tests reject the null hypothesis quite often. On the other hand, for $H(1) : r = 1$ and $\psi_1$ close to 1, all tests are rather conservative in some situations. This is true for both types of trend-adjustment although the estimators proposed by L&S result in slightly better rejection rates than the GLS method. We have also repeated the simulations for the DGPs with $r = 1$ for sample size $T = 200$ and found similar results although the tests are generally less conservative under the null hypothesis in that case.

Table 3. Relative Rejection Frequencies of Test Statistics for DGP (5.1) with Cointegrating Rank $r = 1$, $\psi_2 = 1$, $\delta = 1$, Sample Size $T - 1 = 100$, Nominal Significance Level 0.05.

| Test Statistic | $\psi_1 = 0.95$ | | $\psi_1 = 0.9$ | | $\psi_1 = 0.8$ | | $\psi_1 = 0.7$ | |
|---|---|---|---|---|---|---|---|---|
| | $r_0 = 0$ | $r_0 = 1$ | $r_0 = 0$ | $r_0 = 1$ | $r_0 = 0$ | $r_0 = 1$ | $r_0 = 0$ | $r_0 = 1$ |
| | $\theta = 0$ | | | | | | | |
| $LM^{GLS}$ | 0.044 | 0.007 | 0.075 | 0.016 | 0.253 | 0.037 | 0.535 | 0.041 |
| $LM^{LS}$ | 0.044 | 0.016 | 0.075 | 0.029 | 0.253 | 0.050 | 0.535 | 0.046 |
| $LM_*^{GLS}$ | 0.044 | 0.008 | 0.075 | 0.018 | 0.253 | 0.042 | 0.535 | 0.042 |
| $LM_*^{LS}$ | 0.044 | 0.017 | 0.075 | 0.033 | 0.253 | 0.053 | 0.535 | 0.048 |
| $LR_{trace}^{GLS}$ | 0.065 | 0.007 | 0.104 | 0.018 | 0.313 | 0.047 | 0.613 | 0.047 |
| $LR_{trace}^{LS}$ | 0.065 | 0.016 | 0.104 | 0.030 | 0.313 | 0.057 | 0.613 | 0.053 |
| $LR_{trace}^{J}$ | 0.068 | 0.007 | 0.098 | 0.008 | 0.292 | 0.024 | 0.639 | 0.041 |
| $LR_{trace}^{PC}$ | 0.070 | 0.007 | 0.107 | 0.011 | 0.320 | 0.024 | 0.680 | 0.038 |
| | $\theta = 0.8$ | | | | | | | |
| $LM^{GLS}$ | 0.113 | 0.003 | 0.326 | 0.005 | 0.778 | 0.003 | 0.937 | 0.009 |
| $LM^{LS}$ | 0.113 | 0.017 | 0.326 | 0.024 | 0.778 | 0.020 | 0.937 | 0.014 |
| $LM_*^{GLS}$ | 0.113 | 0.006 | 0.326 | 0.016 | 0.778 | 0.030 | 0.937 | 0.029 |
| $LM_*^{LS}$ | 0.113 | 0.041 | 0.326 | 0.052 | 0.778 | 0.041 | 0.937 | 0.033 |
| $LR_{trace}^{GLS}$ | 0.147 | 0.011 | 0.396 | 0.017 | 0.828 | 0.031 | 0.953 | 0.032 |
| $LR_{trace}^{LS}$ | 0.147 | 0.036 | 0.396 | 0.062 | 0.828 | 0.045 | 0.953 | 0.040 |
| $LR_{trace}^{J}$ | 0.155 | 0.012 | 0.421 | 0.040 | 0.940 | 0.065 | 1.000 | 0.070 |
| $LR_{trace}^{PC}$ | 0.150 | 0.007 | 0.393 | 0.033 | 0.944 | 0.056 | 0.999 | 0.062 |

Table 4. Relative Rejection Frequencies of Test Statistics for DGP (5.1) with Cointegrating Rank $r = 2$, $\psi_2 = 0.5$, $\theta = 0$, $\delta = 0$, Nominal Significance Level 0.05.

| Test Statistic | $\psi_1 = 0.95$ $r_0 = 0$ | $\psi_1 = 0.95$ $r_0 = 1$ | $\psi_1 = 0.9$ $r_0 = 0$ | $\psi_1 = 0.9$ $r_0 = 1$ | $\psi_1 = 0.8$ $r_0 = 0$ | $\psi_1 = 0.8$ $r_0 = 1$ | $\psi_1 = 0.7$ $r_0 = 0$ | $\psi_1 = 0.7$ $r_0 = 1$ |
|---|---|---|---|---|---|---|---|---|
| | $T - 1 = 100$ | | | | | | | |
| $LM^{GLS}$ | 0.881 | 0.090 | 0.919 | 0.227 | 0.972 | 0.612 | 0.993 | 0.842 |
| $LM^{LS}$ | 0.881 | 0.089 | 0.919 | 0.226 | 0.972 | 0.617 | 0.993 | 0.847 |
| $LM_*^{GLS}$ | 0.881 | 0.091 | 0.919 | 0.228 | 0.972 | 0.614 | 0.993 | 0.843 |
| $LM_*^{LS}$ | 0.881 | 0.089 | 0.919 | 0.226 | 0.972 | 0.619 | 0.993 | 0.848 |
| $LR_{trace}^{GLS}$ | 0.916 | 0.102 | 0.936 | 0.245 | 0.982 | 0.640 | 0.995 | 0.855 |
| $LR_{trace}^{LS}$ | 0.916 | 0.100 | 0.936 | 0.248 | 0.982 | 0.640 | 0.995 | 0.860 |
| $LR_{trace}^{J}$ | 0.989 | 0.071 | 0.994 | 0.162 | 1.000 | 0.577 | 1.000 | 0.925 |
| $LR_{trace}^{PC}$ | 0.992 | 0.084 | 0.999 | 0.196 | 1.000 | 0.633 | 1.000 | 0.948 |
| | $T - 1 = 200$ | | | | | | | |
| $LM^{GLS}$ | 0.999 | 0.230 | 1.000 | 0.637 | 1.000 | 0.933 | 1.000 | 0.979 |
| $LM^{LS}$ | 0.999 | 0.227 | 1.000 | 0.637 | 1.000 | 0.930 | 1.000 | 0.980 |
| $LM_*^{GLS}$ | 0.999 | 0.231 | 1.000 | 0.638 | 1.000 | 0.933 | 1.000 | 0.979 |
| $LM_*^{LS}$ | 0.999 | 0.227 | 1.000 | 0.639 | 1.000 | 0.930 | 1.000 | 0.980 |
| $LR_{trace}^{GLS}$ | 0.999 | 0.241 | 1.000 | 0.648 | 1.000 | 0.935 | 1.000 | 0.980 |
| $LR_{trace}^{LS}$ | 0.999 | 0.243 | 1.000 | 0.646 | 1.000 | 0.935 | 1.000 | 0.982 |
| $LR_{trace}^{J}$ | 1.000 | 0.148 | 1.000 | 0.569 | 1.000 | 0.998 | 1.000 | 1.000 |
| $LR_{trace}^{PC}$ | 1.000 | 0.181 | 1.000 | 0.638 | 1.000 | 0.998 | 1.000 | 1.000 |

Table 4 contains results for DGPs with cointegrating rank $r = 2$ ($\psi_2 = 0.5$ and $\psi_1$ varying). Now the processes are stationary and hence the intercept term $\delta$ is set to zero. For $H(1): r = 1$ and both sample sizes the tests with prior trend-adjustment clearly outperform $LR^J_{trace}(r_0)$ and $LR^{PC}_{trace}(r_0)$ when $\psi_1 = 0.95$ or $0.9$, that is, if the alternative is close to the null. This reflects the superior local power of the former tests mentioned in the previous section. Again the method of trend estimation does not matter much. The situation is a little different when $H(0): r = 0$ is tested. For that case the LR tests are more powerful than the LM type tests and $LR^J_{trace}(r_0)$ and $LR^{PC}_{trace}(r_0)$ are generally most powerful.

Hence, the overall conclusion from these simulations is that LR tests with prior trend-adjustment are to be preferred over LM type tests. Which one of the two available trend parameter estimators is used does not make a great difference for the properties of the tests. Thus, one may just as well use the simple GLS method presented in Section 3 rather than the more complicated procedure proposed by L&S. A comparison of the LR type tests for trend-adjusted data with other LR tests which allow for linear trends shows that none of the tests is uniformly superior to all competitors in terms of power. Therefore it seems useful to apply all the available tests simultaneously in samples of the size typical for macroeconomic studies. Although we have not explicitly considered the $LR_{max}$ tests in this simulation study it should be clear that they behave in a similar manner for the DGP used here. In fact, for testing $H(1)$ they are equivalent to the corresponding $LR_{trace}$ tests for the DGP (5.1).

# 6 Conclusions

In this study we have proposed a GLS estimator for the trend parameters of the DGP of a system with cointegrated variables. We have suggested to subtract the trend from the given data first and then perform tests for the cointegrating rank of the system on the basis of the trend-adjusted data. LM and LR type tests based on this idea have been considered. The asymptotic properties of the tests have been derived and are shown to differ from those of the usual LR tests for the cointegrating rank which allow for a linear trend. For some alternatives the asymptotic local power of the new tests is substantially better than that of the standard LR tests. Also, in a simulation study it is found that in some situations the tests based on trend-adjusted data have considerably more power in small samples than standard LR tests which allow for a linear trend. Generally the LR type versions of our new tests outperform the LM type versions. Since in some cases the standard LR tests are superior to the new tests in terms of power, it is recommended to use the old and new tests simultaneously in practice.

# Appendix. Proofs

## A.1 Proof of Theorem 1

We use the notation from Section 3. For ease of exposition we assume for $I(0)$ processes that initial values are taken from the stationary distribution so that these processes are stationary rather than just asymptotically stationary. This does not affect the results. Since all relevant quantities are invariant to normalizations of $\tilde{\alpha}$ and $\tilde{\beta}$, we may assume some kind of normalization and use the following results:

$$\tilde{\alpha} = \alpha + O_p(T^{-1/2}) \quad \text{and} \quad \tilde{\beta} = \beta + O_p(T^{-1}) \tag{A.1}$$

Moreover, the estimators $\tilde{\Gamma}_i$ and $\tilde{\Omega}$ are consistent (see Johansen (1995)).

In the following we show that the results stated in Theorem 1 hold if $\beta$ and $\beta_\perp$ are replaced by $\tilde{\beta}$ and $\tilde{\beta}_\perp$. Then the theorem follows from $(A.1)$. Hence, we consider the "parameters" $\underline{\delta} = \tilde{\beta}'\mu_0$, $\underline{\delta}_* = \tilde{\beta}_\perp'\mu_0$, $\underline{\tau} = \tilde{\beta}'\mu_1$ and $\underline{\tau}_* = \tilde{\beta}_\perp'\mu_1$. The corresponding "real" parameters are $\delta = \beta'\mu_0$, $\delta_* = \beta_\perp'\mu_0$, $\tau = \beta'\mu_1$ and $\tau_* = \beta_\perp'\mu_1$. The definition of $\tilde{G}_t$ implies that $\tilde{G}_t\tilde{\beta}_\perp(\tilde{\beta}_\perp'\tilde{\beta}_\perp)^{-1} = 0,\ t \geq p+1$. Using the definitions of the variables in Section 3 it is therefore straightforward to see that the LS estimator of $\underline{\delta}_*$ obtained from (3.4) and denoted by $\hat{\underline{\delta}}_*$ has no effect on the asymptotic properties of the other estimators obtained from (3.4). L&S further show that the moment matrix related to the LS estimator of $\hat{\underline{\delta}}_*$ is asymptotically nonsingular. Hence, it follows that

$$\hat{\underline{\delta}}_* = \tilde{\beta}_\perp'\hat{\mu}_0 = \tilde{\beta}_\perp'\mu_0 + O_p(1) = \delta_* + O_p(1) \tag{A.2}$$

which implies (3.6). Thus, we can concentrate on the estimation of $\underline{\delta}$, $\underline{\tau}$ and $\underline{\tau}_*$ and ignore the first $p$ observations in (3.4). From the definitions it follows that we have, for $t = p+1, \ldots, T$,

$$\tilde{d}_t := \tilde{Q}'\tilde{A}(L)y_t = -\tilde{Q}'\tilde{\alpha}\underline{\delta} + \tilde{Q}'(\tilde{\Psi}\tilde{\beta}(\tilde{\beta}'\tilde{\beta})^{-1} - (t-1)\tilde{\alpha})\underline{\tau} + \tilde{Q}'\tilde{\Psi}\tilde{\beta}_\perp(\tilde{\beta}_\perp'\tilde{\beta}_\perp)^{-1}\underline{\tau}_* + \eta_t. \tag{A.3}$$

The derivation of this equation from (3.4) is based on calculations similar to those used to obtain equation (5.8) from (5.7) in L&S.

Now consider the LS estimators of $\underline{\delta}$, $\underline{\tau}$ and $\underline{\tau}_*$ obtained from $(A.3)$. We first prove the following result for the estimators $\hat{\underline{\delta}}$ and $\hat{\underline{\tau}}$.

**Lemma A.1**

Under the conditions of Theorem 1,

$$\hat{\underline{\delta}} = \underline{\delta} + O_p(T^{-1/2}) \quad \text{and} \quad \hat{\underline{\tau}} = \underline{\tau} + O_p(T^{-3/2}).$$

**Proof:** We have $\eta_t = \tilde{Q}' n_t$, where

$$\begin{aligned} n_t &= \varepsilon_t - (\tilde{\alpha}\tilde{\beta}' - \alpha\beta')x_{t-1} - \sum_{j=1}^{p-1}(\tilde{\Gamma}_j - \Gamma_j)\Delta x_{t-j} \\ &= \varepsilon_t - \tilde{\alpha}(\tilde{\beta} - \beta)' x_{t-1} - (\tilde{\alpha} - \alpha)\beta' x_{t-1} - \sum_{j=1}^{p-1}(\tilde{\Gamma}_j - \Gamma_j)\Delta x_{t-j}. \end{aligned} \tag{A.4}$$

Using this expression of $n_t$ it is straightforward to study the limiting behavior of the cross products between the components of the error term $\eta_t$ and regressors in $(A.3)$. First note that, using (3.3) and the identity $\eta_t = \tilde{Q}' n_t$,

$$\begin{aligned} -\tilde{\alpha}'\tilde{Q}T^{-1/2}\sum_{t=p+1}^{T}\eta_t &= -\tilde{\alpha}'\tilde{\Omega}^{-1}T^{-1/2}\sum_{t=p+1}^{T} n_t \\ &= -\tilde{\alpha}'\tilde{\Omega}^{-1}T^{-1/2}\sum_{t=p+1}^{T}(\varepsilon_t - \tilde{\alpha}(\tilde{\beta} - \beta)' x_{t-1}) + o_p(1), \end{aligned}$$

where the latter equality follows from $(A.4)$ because $\beta' x_t$ and $\Delta x_t$ are zero mean stationary processes and the estimators $\tilde{\alpha}$ and $\tilde{\Gamma}_j$ are consistent. Writing

$$(\tilde{\beta} - \beta)' x_{t-1} = (\tilde{\beta} - \beta)'\beta(\beta'\beta)^{-1}\beta' x_{t-1} + (\tilde{\beta} - \beta)'\beta_\perp(\beta_\perp'\beta_\perp)^{-1}v_{t-1}$$

with $v_t = \beta_\perp' x_t$, we can further see that

$$\begin{aligned} -\tilde{\alpha}'\tilde{Q}T^{-1/2}\sum_{t=p+1}^{T}\eta_t &= -\alpha'\Omega^{-1}T^{-1/2}\sum_{t=p+1}^{T}\varepsilon_t \\ &\quad +\alpha'\Omega^{-1}\alpha(\tilde{\beta} - \beta)'\beta_\perp(\beta_\perp'\beta_\perp)^{-1}T^{-1/2}\sum_{t=p+1}^{T} v_{t-1} + o_p(1) \\ &= O_p(1). \end{aligned} \tag{A.5}$$

Next consider the cross products between the components of $\eta_t$ and the second set of regressors in $(A.3)$. It is clear that the limiting behaviour of these quantities is dominated by the trend term of the regressor so that we have

$$T^{-3/2}\sum_{t=p+1}^{T}\left[(\tilde{\beta}'\tilde{\beta})^{-1}\tilde{\beta}'\tilde{\Psi}' - (t-1)\tilde{\alpha}'\right]\tilde{Q}\eta_t$$

$$= -\tilde{\alpha}'\tilde{Q}T^{-3/2}\sum_{t=p+1}^{T}(t-1)\eta_t + o_p(1)$$

$$= -\alpha'\Omega^{-1}T^{-3/2}\sum_{t=p+1}^{T}(t-1)\varepsilon_t + \alpha'\Omega^{-1}\alpha(\tilde{\beta}-\beta)'\beta_\perp(\beta_\perp'\beta_\perp)^{-1}T^{-3/2}\sum_{t=p+1}^{T}(t-1)v_{t-1} + o_p(1)$$

$$= O_p(1). \qquad (A.6)$$

Here the second equality is obtained in the same way as in $(A.5)$. Finally, note that similar arguments can be used to obtain a representation for the cross products between the components of $\eta_t$ and the third set of regressors in $(A.3)$. However, at this stage it suffices to mention that

$$(\tilde{\beta}_\perp'\tilde{\beta}_\perp)^{-1}\tilde{\beta}_\perp'\tilde{\Psi}'\tilde{Q}T^{-1/2}\sum_{t=p+1}^{T}\eta_t = O_p(1). \qquad (A.7)$$

Now consider the (appropriately standardized) moment matrix related to the LS estimation of $(A.3)$. In the same way as above, we can clearly ignore the term $\tilde{Q}'\tilde{\Psi}\tilde{\beta}(\tilde{\beta}'\tilde{\beta})^{-1}$ in the second set of regressors. For notational convenience and without loss of generality we study this moment matrix for $t = 1, \ldots, T$. Define

$$c_T = [1 \quad c_{1T}]' \quad \text{and} \quad C_T = \begin{bmatrix} 1 & c_{1T} \\ c_{1T} & c_{2T} \end{bmatrix},$$

where

$$c_{1T} = T^{-2}\sum_{t=1}^{T}(t-1) \quad \text{and} \quad c_{2T} = T^{-3}\sum_{t=1}^{T}(t-1)^2.$$

Clearly, $c_{1T} = \frac{1}{2} + o(1)$ and $c_{2T} = \frac{1}{3} + o(1)$. Below we shall also use the obvious results $c_T'C_T^{-1} = [1 \quad 0]$ and $c_T'C_T^{-1}c_T = 1$. The moment matrix we wish to study can now be written as

$$\sum_{t=1}^{T}\begin{bmatrix} -T^{-1/2}\tilde{\alpha}'\tilde{Q} \\ -T^{-3/2}(t-1)\tilde{\alpha}'\tilde{Q} \\ T^{-1/2}(\tilde{\beta}_\perp'\tilde{\beta}_\perp)^{-1}\tilde{\beta}_\perp'\tilde{\Psi}'\tilde{Q} \end{bmatrix}\left[-T^{-1/2}\tilde{Q}'\tilde{\alpha} : -T^{-3/2}(t-1)\tilde{Q}'\tilde{\alpha} : T^{-1/2}\tilde{Q}'\tilde{\Psi}\tilde{\beta}_\perp(\tilde{\beta}_\perp'\tilde{\beta}_\perp)^{-1}\right]$$

$$= \begin{bmatrix} C_T \otimes \tilde{A}_{11} & c_T \otimes \tilde{A}_{12} \\ c_T' \otimes \tilde{A}_{21} & \tilde{A}_{22} \end{bmatrix}. \qquad (A.8)$$

Here we have, using (3.3),

$$\tilde{A}_{11} = \tilde{\alpha}'\tilde{\Omega}^{-1}\tilde{\alpha}$$

$$\tilde{A}_{12} = \tilde{A}_{21}' = -\tilde{\alpha}'\tilde{\Omega}^{-1}\tilde{\Psi}\tilde{\beta}_\perp(\tilde{\beta}_\perp'\tilde{\beta}_\perp)^{-1}$$

and

$$\tilde{A}_{22} = (\tilde{\beta}_\perp'\tilde{\beta}_\perp)^{-1}\tilde{\beta}_\perp'\tilde{\Psi}'\tilde{\Omega}^{-1}\tilde{\Psi}\tilde{\beta}_\perp(\tilde{\beta}_\perp'\tilde{\beta}_\perp)^{-1} = \tilde{A}_{21}\tilde{A}_{11}^{-1}\tilde{A}_{12} + \tilde{B}^{-1}$$

where

$$\tilde{B} = \tilde{\beta}_\perp'\tilde{\beta}_\perp(\tilde{\alpha}_\perp'\tilde{\Psi}\tilde{\beta}_\perp)^{-1}\tilde{\alpha}_\perp'\tilde{\Omega}\tilde{\alpha}_\perp(\tilde{\beta}_\perp'\tilde{\Psi}'\tilde{\alpha}_\perp)^{-1}\tilde{\beta}_\perp'\tilde{\beta}_\perp = \tilde{\beta}_\perp'\tilde{C}\tilde{\Omega}\tilde{C}'\tilde{\beta}_\perp$$

with $\tilde{C} = \tilde{\beta}_\perp(\tilde{\alpha}_\perp'\tilde{\Psi}\tilde{\beta}_\perp)^{-1}\tilde{\alpha}_\perp'$, a sample analog of the matrix $C$. The latter expression of $\tilde{A}_{22}$ is obtained by using the latter equality in (3.3) and direct calculation. It is straightforward to check that the matrix on the r.h.s. of $(A.8)$ converges in probability to a nonsingular limit. This fact together with $(A.5) - (A.7)$ imply Lemma A.1. □

For the LS estimator $\hat{\underline{\tau}}_*$ of $\underline{\tau}_*$ obtained from $(A.3)$ we give the asymptotic distribution in the next lemma.

**Lemma A.2**

Under the conditions of Theorem 1,

$$\sqrt{T}(\hat{\underline{\tau}}_* - \underline{\tau}_*) \xrightarrow{d} N(0, \beta_\perp' C\Omega C'\beta_\perp).$$

**Proof:** To obtain the asymptotic distribution of $\hat{\tau}_*$, we have to calculate the lower part of the inverse of the matrix on the r.h.s. of $(A.8)$ and multiply the vectors in $(A.5) - (A.7)$ by it. Using the well-known formula for the partitioned inverse and the results $c_T' C_T^{-1} = [1 \quad 0]$ and $c_T' C_T^{-1} c_T = 1$ one readily obtains

$$\begin{bmatrix} C_T \otimes \tilde{A}_{11} & c_T \otimes \tilde{A}_{12} \\ c_T' \otimes \tilde{A}_{21} & \tilde{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} * & * & * \\ -\tilde{B}\tilde{A}_{21}\tilde{A}_{11}^{-1} & 0 & \tilde{B} \end{bmatrix},$$

where the blocks denoted by "*" are not needed and the partition on the r.h.s. conforms to the three estimators $\hat{\underline{\delta}}$, $\hat{\underline{\tau}}$ and $\hat{\underline{\tau}}_*$. Thus, it follows that the estimator $\hat{\underline{\tau}}_*$ satisfies

$$\begin{aligned} \hat{\underline{\tau}}_* - \underline{\tau}_* &= \tilde{B}\tilde{A}_{21}\tilde{A}_{11}^{-1}\tilde{\alpha}'\tilde{Q}T^{-1}\sum_{t=1}^{T}\eta_t + \tilde{B}(\tilde{\beta}_\perp'\tilde{\beta}_\perp)^{-1}\tilde{\beta}_\perp'\tilde{\Psi}'\tilde{Q}T^{-1}\sum_{t=1}^{T}\eta_t + o_p(T^{-1/2}) \\ &= \tilde{B}\left(\tilde{A}_{21}\tilde{A}_{11}^{-1}\tilde{\alpha}' + (\tilde{\beta}_\perp'\tilde{\beta}_\perp)^{-1}\tilde{\beta}_\perp'\tilde{\Psi}'\right)\tilde{\Omega}^{-1}T^{-1}\sum_{t=1}^{T}n_t + o_p(T^{-1/2}) \end{aligned}$$

where we make use of (3.3), $(A.5)$ and $(A.7)$ and the term $o_p(T^{-1/2})$ is only due to ignoring the estimation of $\underline{\delta}_*$ and starting the summation from $t=1$ instead of $t=p+1$. By the definitions and the latter equality in (3.3),

$$\begin{aligned}
\tilde{B}\left[\tilde{A}_{21}\tilde{A}_{11}^{-1}\tilde{\alpha}' + (\tilde{\beta}_\perp'\tilde{\beta}_\perp)^{-1}\tilde{\beta}_\perp'\tilde{\Psi}'\right] &= -\tilde{B}\left[(\tilde{\beta}_\perp'\tilde{\beta}_\perp)^{-1}\tilde{\beta}_\perp'\tilde{\Psi}'\tilde{\Omega}^{-1}\tilde{\alpha}(\tilde{\alpha}'\tilde{\Omega}^{-1}\tilde{\alpha})^{-1}\tilde{\alpha}' - (\tilde{\beta}_\perp'\tilde{\beta}_\perp)^{-1}\tilde{\beta}_\perp'\tilde{\Psi}'\right] \\
&= \tilde{B}(\tilde{\beta}_\perp'\tilde{\beta}_\perp)^{-1}\tilde{\beta}_\perp'\tilde{\Psi}'\tilde{\alpha}_\perp(\tilde{\alpha}_\perp'\tilde{\Omega}\tilde{\alpha}_\perp)^{-1}\tilde{\alpha}_\perp'\tilde{\Omega} \\
&= \tilde{\beta}_\perp'\tilde{\beta}_\perp(\tilde{\alpha}_\perp'\tilde{\Psi}\tilde{\beta}_\perp)^{-1}\tilde{\alpha}_\perp'\tilde{\Omega} \\
&= \tilde{\beta}_\perp'\tilde{C}\tilde{\Omega}.
\end{aligned}$$

Thus, it follows that

$$\begin{aligned}
T^{1/2}(\hat{\underline{\tau}}_* - \underline{\tau}_*) &= \tilde{\beta}_\perp'\tilde{C}T^{-1/2}\sum_{t=1}^{T} n_t + o_p(1) \\
&= \beta_\perp' C T^{-1/2}\sum_{t=1}^{T}\varepsilon_t + o_p(1)
\end{aligned}$$

where the latter equality is obtained by using the representation of $e_t$ given in $(A.4)$ and arguments similar to those in $(A.5)$. Thus, since by $(A.1)$ we can assume that $\tilde{\beta}_\perp = \beta_\perp + O_p(T^{-1})$, the lemma follows. □

Recall that $\tilde{\beta} = \beta + O_p(T^{-1})$ can be assumed. Thus, from Lemmas A.1 and A.2 we get $\tilde{\mu}_1 = O_p(1)$ and, hence, (3.8) by noting that

$$\sqrt{T}(\hat{\underline{\tau}}_* - \underline{\tau}_*) = \sqrt{T}\beta_\perp'(\tilde{\mu}_1 - \mu_1) + \sqrt{T}(\tilde{\beta}_\perp' - \beta_\perp)'(\tilde{\mu}_1 - \mu_1) = \sqrt{T}\beta_\perp'(\tilde{\mu}_1 - \mu_1) + o_p(1).$$

Thus, it remains to prove (3.5) and (3.7).

By $(A.2)$ and Lemma A.1, $\tilde{\mu}_0 = O_p(1)$ so that

$$\begin{aligned}
\hat{\underline{\delta}} - \underline{\delta} &= \beta'(\tilde{\mu}_0 - \mu_0) + (\tilde{\beta} - \beta)'(\tilde{\mu}_0 - \mu_0) \\
&= \beta'(\tilde{\mu}_0 - \mu_0) + O_p(T^{-1}).
\end{aligned} \tag{A.9}$$

Hence, the limiting distribution of $\tilde{\mu}_0$ in the direction of $\beta$ is the same as the limiting distribution of $\hat{\underline{\delta}}$ so that, in particular, (3.5) follows. The argument used in $(A.9)$ can be repeated for $\tilde{\mu}_1$ to show that (3.7) holds. Thereby Theorem 1 is proven.

## A.2 Proof of Theorem 3

The idea of the proof is to adopt the arguments used by Johansen (1995, pp. 158 - 161) to derive the asymptotic properties of the test statistics. For this purpose we give some intermediate results.

**Lemma A.3**

(i) $(T-p)^{-1}\beta' \sum_{t=p+1}^{T} \tilde{x}_{t-1}\tilde{x}'_{t-1}\beta \xrightarrow{p} E(u_t u'_t)$

(ii) $(T-p)^{-2}\beta'_\perp \sum_{t=p+1}^{T} \tilde{x}_{t-1}\tilde{x}'_{t-1}\beta_\perp \xrightarrow{d} \beta'_\perp C\Omega^{1/2}\int_0^1 \mathbf{B}_*(s)\mathbf{B}_*(s)'\mathrm{d}s\Omega^{1/2}C'\beta_\perp$

(iii) $(T-p)^{-1} \sum_{t=p+1}^{T} \Delta\tilde{x}_{t-j}\Delta\tilde{x}'_{t-k} \xrightarrow{p} E(\Delta x_{t-j}\Delta x'_{t-k}) \qquad (j,k=0,\ldots,p-1)$

(iv) $(T-p)^{-1}\beta' \sum_{t=p+1}^{T} \tilde{x}_{t-1}\Delta\tilde{x}'_{t-k} \xrightarrow{p} E(u_{t-1}\Delta x'_{t-k}) \qquad (k=0,\ldots,p-1)$

(v) $(T-p)^{-1} \sum_{t=p+1}^{T} \Delta\tilde{x}_{t-j}\tilde{x}'_{t-1} = O_p(1) \qquad (j=0,\ldots,p-1)$

(vi) $(T-p)^{-1}\beta' \sum_{t=p+1}^{T} \tilde{x}_{t-1}\tilde{x}'_{t-1} = O_p(1).$

**Proof:** First note that $\beta' x_t$ and $\beta'_\perp \Delta x_t$ are stationary processes with zero mean and that

$$\tilde{x}_t = x_t - (\tilde{\mu}_0 - \mu_0) - (\tilde{\mu}_1 - \mu_1)t. \qquad (A.10)$$

The first result follows straightforwardly from this identity, Theorem 1 ((3.5), (3.7)) and well-known limit theorems. The second result can be obtained from the proof of Theorem 5.1 of L&S [see the derivation of $(A.14)$ and $(A.15)$]. Finally, the last four results are again straightforward consequences of $(A.10)$, Theorem 1, and well-known limit theorems. □

Lemma A.3 can be used to analyse sample moments of the regressors in (4.6). The next lemma provides similar results for the sample moments of the regressors and the error term in (4.6).

**Lemma A.4**

(i) $(T-p)^{-1/2}\beta' \sum_{t=p+1}^{T} \tilde{x}_{t-1}e'_t = O_p(1)$

(ii) $(T-p)^{-1}\beta_\perp' \sum_{t=p+1}^{T} \tilde{x}_{t-1} e_t' = O_p(1)$

(iii) $(T-p)^{-1/2} \sum_{t=p+1}^{T} \Delta\tilde{x}_{t-j} e_t' = O_p(1)$

**Proof:** From (2.3) and (4.6) it follows that

$$\begin{aligned} e_t &= \varepsilon_t - \alpha\beta'(\tilde{x}_{t-1} - x_{t-1}) + \Delta\tilde{x}_t - \Delta x_t - \textstyle\sum_{i=1}^{p-1} \Gamma_i(\Delta\tilde{x}_{t-i} - \Delta x_{t-i}) \\ &= \varepsilon_t - \alpha\beta'(\tilde{\mu}_0 - \mu_0) - \alpha\beta'(\tilde{\mu}_1 - \mu_1)(t-1) - \Psi(\tilde{\mu}_1 - \mu_1). \end{aligned}$$

The proof of Lemma A.4 is obtained by using this expression and ($A$.10) and applying Theorem 1 in conjunction with well-known limit theorems. Details are straightforward but somewhat tedious and, therefore, omitted. □

Now suppose that $\alpha, \beta$ and the residual covariance matrix are estimated from (4.6) in precisely the way described by Johansen (1988). The estimators will be denoted by $\bar{\alpha}$, $\bar{\beta}$ and $\bar{\Omega}$, respectively. For these estimators the following properties can be shown.

**Lemma A.5**

Consider the normalized estimators $\bar{\beta}_\gamma = \bar{\beta}(\gamma'\bar{\beta})^{-1}$ and $\bar{\alpha}_\gamma = \bar{\alpha}\bar{\beta}'\gamma$ where $\gamma' = (\beta'\beta)^{-1}\beta'$. Then, $\bar{\beta}_\gamma = \beta + O_p(T^{-1})$, $\bar{\alpha}_\gamma = \alpha + O_p(T^{-1/2})$ and $\bar{\Omega} = \Omega + O_p(T^{-1/2})$.

**Proof:** The proof can be obtained by using Lemmas A.3 and A.4 in conjunction with arguments used in the proofs of Lemmas 13.1 and 13.2 of Johansen (1995). At this point it may be worth noting that the results of Lemmas A.3 and A.4 are in some respects similar to those obtained for the infeasible model (2.3). Specifically, the results (i), (iii) and (iv) of Lemma A.3 which are concerned with the "stationary" series $\beta'\tilde{x}_{t-1}$ and $\Delta\tilde{x}_{t-j}$, are exactly the same as their counterparts obtained with $\tilde{x}_t$ replaced by $x_t$ while the remaining results of Lemma A.3, which involve the "nonstationary" series $\beta_\perp'\tilde{x}_{t-1}$, show that the rates of convergence are the same as in the case where $\tilde{x}_{t-1}$ is replaced by its unobservable counterpart $x_{t-1}$. Similarly, the rates of convergence in Lemma A.4 are the same as those obtained with $\tilde{x}_t$ and $e_t$ replaced by $x_t$ and $\varepsilon_t$, respectively. Keeping these facts in mind it is straightforward to obtain the proof by following the above mentioned proofs of Johansen (1995). We shall only give an outline of the main steps.

First note that the result $\bar{\beta}_\gamma = \beta + o_p(T^{-1/2})$ can be proved by making appropriate modifications to the proof of Lemma 13.1 of Johansen (1995). In place of Johansen's (1995) equation (13.6) we can use a similar equation with $A_T = [\beta : T^{-1/2}\beta_\perp]$ and the involved moment matrices replaced by analogs obtained from the auxiliary model (4.6). By Lemma A.3, the asymptotic behavior of these matrices is entirely similar to those of their counterparts in Johansen's (1995) equation (13.6). The only exception is the form of the weak limit in Lemma A.3 (ii) but this has no effect on the consistency proof. As the next step, we can prove the consistency of the estimators $\bar{\alpha}_\gamma$ and $\bar{\Omega}$ by following the corresponding consistency proof in Johansen's (1995) Lemma 13.1. In our case Johansen's (1995) matrix $B_T$ should be defined as $B_T = \beta_\perp$. The next step is to establish the stated orders of consistency. First we can write the first order conditions for $\bar{\alpha}_\gamma$ and $\bar{\beta}_\gamma$ by modifying Johansen's (1995) equations (13.8) and (13.9) in an obvious way after which the proof proceeds in the same way as in Johansen (1995, pp. 182-183) except that the relevant convergence results are obtained from Lemmas A.3 and A.4.

The result $\bar{\Omega} = \Omega + O_p(T^{-1/2})$ is not explicitly considered in Johansen (1995) but it can be obtained in a straightforward manner from the order results for $\bar{\alpha}_\gamma$ and $\bar{\beta}_\gamma$. □

Lemma A.5 implies that the same consistency results also hold for other normalizations (see Johansen (1995, p. 184)). In what follows we again assume that some kind of normalization has been applied to $\bar{\alpha}$ and $\bar{\beta}$.

**Lemma A.6**

Defining $\bar{u}_t = \bar{\beta}'\tilde{x}_t$, $\bar{v}_t = \bar{\beta}_\perp'\tilde{x}_t$, $\Delta\bar{X}^{*\prime}_{t-1} = [\bar{u}'_{t-1}, \Delta x'_{t-1}, \ldots, \Delta x'_{t-p+1}]$ and $e_{*t} = e_t - \alpha(\bar{\beta} - \beta)'\tilde{x}_{t-1}$ we have

$$(T-p)^{-2}\sum_{t=p+1}^{T}\bar{v}_{t-1}\bar{v}'_{t-1} \xrightarrow{d} \beta_\perp' C\Omega^{1/2}\int_0^1 \mathbf{B}_*(s)\mathbf{B}_*(s)'\mathrm{d}s\Omega^{1/2}C'\beta_\perp,$$

$$(T-p)^{-3/2}\sum_{t=p+1}^{T}\bar{v}_{t-1}\Delta\bar{X}^{*\prime}_{t-1} = o_p(1)$$

and

$$(T-p)^{-1}\sum_{t=p+1}^{T}\bar{v}_{t-1}e'_{*t}\bar{\alpha}_\perp \xrightarrow{d} \beta_\perp' C\Omega^{1/2}\int_0^1 \mathbf{B}_*(s)\mathrm{d}\mathbf{B}_*(s)'\Omega^{1/2}\alpha_\perp.$$

**Proof:** The first two results of the lemma readily follow from the consistency of the estimators $\bar{\beta}$ and $\bar{\beta}_\perp$ and Lemma A.3. To prove the last assertion of the lemma we note that using the definition of $e_{*t}$, the consistency of the estimators $\bar{\alpha}_\perp$ and $\bar{\beta}_\perp$ and Lemma A.3 it is first straightforward to show that

$$\begin{aligned}(T-p)^{-1}\sum_{t=p+1}^{T}\bar{v}_{t-1}e_{*t}'\bar{\alpha}_\perp &= (T-p)^{-1}\sum_{t=p+1}^{T}\bar{v}_{t-1}e_t'\bar{\alpha}_\perp + o_p(1)\\ &= (T-p)^{-1}\sum_{t=p+1}^{T}\bar{\beta}_\perp' x_{t-1}e_t'\bar{\alpha}_\perp\\ &\quad -\bar{\beta}_\perp'(\tilde{\mu}_0-\mu_0)(T-p)^{-1}\sum_{t=p+1}^{T}e_t'\bar{\alpha}_\perp\\ &\quad -\bar{\beta}_\perp'(\tilde{\mu}_1-\mu_1)(T-p)^{-1}\sum_{t=p+1}^{T}(t-1)e_t'\bar{\alpha}_\perp + o_p(1),\end{aligned} \tag{A.11}$$

where the second equality follows from the identity $\bar{v}_{t-1} = \bar{\beta}_\perp'\tilde{x}_{t-1}$ and $(A.10)$. Next note that, by the definition of $e_t$,

$$\begin{aligned}(T-p)^{-1}\sum_{t=p+1}^{T}\bar{\beta}_\perp' x_{t-1}e_t'\bar{\alpha}_\perp &= (T-p)^{-1}\sum_{t=p+1}^{T}\bar{\beta}_\perp' x_{t-1}\varepsilon_t'\bar{\alpha}_\perp\\ &\quad -(T-p)^{-1}\sum_{t=p+1}^{T}\bar{\beta}_\perp' x_{t-1}(\tilde{\mu}_0-\mu_0)'\beta\alpha'\bar{\alpha}_\perp\\ &\quad -(T-p)^{-1}\sum_{t=p+1}^{T}\bar{\beta}_\perp' x_{t-1}(t-1)(\tilde{\mu}_1-\mu_1)'\beta\alpha'\bar{\alpha}_\perp\\ &\quad -(T-p)^{-1}\sum_{t=p+1}^{T}\bar{\beta}_\perp' x_{t-1}(\tilde{\mu}_1-\mu_1)'\Psi'\bar{\alpha}_\perp\\ &= (T-p)^{-1}\sum_{t=p+1}^{T}\beta_\perp' x_{t-1}\varepsilon_t'\alpha_\perp\\ &\quad -(T-p)^{-1}\sum_{t=p+1}^{T}\beta_\perp' x_{t-1}(\tilde{\mu}_1-\mu_1)'\beta_\perp(\beta_\perp'\beta_\perp)^{-1}\beta_\perp'\Psi'\alpha_\perp\\ &\quad +o_p(1)\end{aligned}$$

where the latter equality is an immediate consequence of Theorem 1 and the consistency of $\bar{\alpha}_\perp$ and $\bar{\beta}_\perp$. Similarly,

$$\bar{\beta}_\perp'(\tilde{\mu}_0-\mu_0)(T-p)^{-1}\sum_{t=p+1}^{T}e_t'\bar{\alpha}_\perp = o_p(1)$$

and

$$\begin{aligned}&\bar{\beta}_\perp'(\tilde{\mu}_1-\mu_1)(T-p)^{-1}\sum_{t=p+1}^{T}(t-1)e_t'\bar{\alpha}_\perp\\ &= \beta_\perp'(\tilde{\mu}_1-\mu_1)(T-p)^{-1}\sum_{t=p+1}^{T}(t-1)\varepsilon_t'\alpha_\perp\\ &\quad -\beta_\perp'(\tilde{\mu}_1-\mu_1)(T-p)^{-1}\sum_{t=p+1}^{T}(t-1)(\tilde{\mu}_1-\mu_1)'\beta_\perp(\beta_\perp'\beta_\perp)^{-1}\beta_\perp'\Psi'\alpha_\perp + o_p(1).\end{aligned}$$

Combining the above results with $(A.11)$ and using the notation $v_t = \beta_\perp' x_t$ yields

$$
\begin{aligned}
&(T-p)^{-1} \sum_{t=p+1}^{T} \bar{v}_{t-1} e_{*t}' \bar{\alpha}_\perp \\
&= (T-p)^{-1} \sum_{t=p+1}^{T} v_{t-1} \varepsilon_t' \alpha_\perp \\
&\quad -(T-p)^{-1} \sum_{t=p+1}^{T} v_{t-1} (\tilde{\mu}_1 - \mu_1)' \beta_\perp (\beta_\perp' \beta_\perp)^{-1} \beta_\perp' \Psi' \alpha_\perp \\
&\quad -\beta_\perp' (\tilde{\mu}_1 - \mu_1)(T-p)^{-1} \sum_{t=p+1}^{T} (t-1) \varepsilon_t' \alpha_\perp \\
&\quad +\beta_\perp' (\tilde{\mu}_1 - \mu_1)(T-p)^{-1} \sum_{t=p+1}^{T} (t-1)(\tilde{\mu}_1 - \mu_1)' \beta_\perp (\beta_\perp' \beta_\perp)^{-1} \beta_\perp' \Psi' \alpha_\perp + o_p(1) \\
&= (T-p)^{-1} \sum_{t=p+1}^{T} [v_{t-1} - \beta_\perp' (\tilde{\mu}_1 - \mu_1)(t-1)][\varepsilon_t' \alpha_\perp C' \beta_\perp - (\tilde{\mu}_1 - \mu_1)' \beta_\perp](\beta_\perp' \beta_\perp)^{-1} \beta_\perp' \Psi' \alpha_\perp \\
&\quad + o_p(1).
\end{aligned}
$$

The last expression is the same as $(A.32)$ of L&S expect for the factor $(\beta_\perp'\beta_\perp)^{-1}\beta_\perp'\Psi'\alpha_\perp$, so Lemma A.6 follows from $(A.33)$ of L&S and the definition of the matrix $C$. This completes the proof of Lemma A.6. $\square$

Now we can prove Theorem 3. First consider the test statistic $LR_{trace}^{GLS}(r_0)$ and the auxiliary regression model

$$
\bar{\alpha}_\perp' \Delta \tilde{x}_t = \kappa^* \bar{u}_{t-1} + \rho^* \bar{v}_{t-1} + \sum_{j=1}^{p-1} \Gamma_j^* \Delta \tilde{x}_{t-j} + \bar{\alpha}_\perp' e_{*t}, \qquad t = p+1, \ldots, T, \qquad (A.12)
$$

where $e_{*t} = e_t - \alpha(\bar{\beta} - \beta)' \tilde{x}_{t-1}$ as before. From Saikkonen & Lütkepohl (1997) we can conclude that the test statistic $LR_{trace}^{GLS}(r_0)$ can be obtained from $(A.12)$ as the conventional LR test statistic of the multivariate linear model for the null hypothesis $\rho^* = 0$. This LR test statistic is asymptotically equivalent to the corresponding Wald test statistic. Using Lemmas A.3 - A.6 it can further be shown that an asymptotically equivalent Wald statistic is obtained by deleting the regressors $\bar{u}_{t-1}$ and $\Delta\tilde{x}_{t-j}$ $(j = 1, \ldots, p-1)$ from $(A.12)$. Hence, we have

$$
LR_{trace}^{GLS}(r_0) = \operatorname{tr} \left\{ (\bar{\alpha}_\perp' \bar{\Omega} \bar{\alpha}_\perp)^{-1} \bar{\alpha}_\perp' \sum_{t=p+1}^{T} e_{*t} \bar{v}_{t-1}' \left( \sum_{t=p+1}^{T} \bar{v}_{t-1} \bar{v}_{t-1}' \right)^{-1} \sum_{t=p+1}^{T} \bar{v}_{t-1} e_{*t}' \bar{\alpha}_\perp \right\} + o_p(1).
$$

Recalling that $C = \beta_\perp(\alpha_\perp'\Psi\beta_\perp)^{-1}\alpha_\perp'$ and following the arguments in L&S, the stated result follows from this representation and Lemmas A.5 and A.6.

The above proof is based on the approach in L&S which does not require the derivation of the joint limiting distribution of the eigenvalues $\bar{\lambda}_i$ $(i = r+1, \ldots, n)$. Therefore this approach cannot be used to derive the limiting distribution of the test statistic $LR^{GLS}_{max}(r_0)$. However, using the results of Lemmas A.3 - A.6 and proceeding as in Johansen (1995, pp. 158 - 161) or Saikkonen & Luukkonen (1997), for example, it is straightforward to obtain the limiting distribution of the test statistic $LR^{GLS}_{max}(r_0)$ as well. Details are omitted.

# References


Engle, R.F. & C.W.J. Granger (1987), Co-integration and error correction: Representation, estimation and testing, *Econometrica*, 55, 251-276.

Johansen, S. (1988), Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control*, 12, 231-254.

Johansen, S. (1991), Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica*, 59, 1551-1580.

Johansen, S. (1992), Determination of cointegration rank in the presence of a linear trend, *Oxford Bulletin of Economics and Statistics*, 54, 383-397.

Johansen, S. (1995), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford, Oxford University Press.

Lütkepohl, H. & P. Saikkonen (1997), Testing for the cointegrating rank of a VAR process with a time trend, Discussion Paper, SFB 373, Humboldt Universität Berlin.

Perron, P. & J.Y. Campbell (1993), A note on Johansen's cointegration procedure when trends are present, *Empirical Economics*, 18, 777-789.

Rahbek, A.C. (1994), The power of some multivariate cointegration tests, Discussion Paper, Institute of Mathematical Statistics, University of Copenhagen.

Saikkonen, P. & H. Lütkepohl (1997), Local power of likelihood ratio tests for the cointegrating rank of a VAR process, Discussion Paper No. 58, SFB 373, Humboldt Universität Berlin.

Saikkonen, P. & R. Luukkonen (1997), Testing cointegration in infinite order vector autoregressive processes, *Journal of Econometrics*, forthcoming.

Stock, J.H. & M.W. Watson (1988), Testing for common trends, *Journal of the American Statistical Association*, 83, 1097-1107.

Toda, H.Y. (1994), Finite sample properties of likelihood ratio tests for cointagrating ranks when linear trends are present, *Review of Economics and Statistics*, 76, 66 - 79.

Toda, H.Y. (1995), Finite sample performance of likelihood ratio tests for cointagrating ranks in vector autoregressions, *Econometric Theory*, 11, 1015 - 1032.